



(U) How Do You Fix a Damaged PDF?

FROM: SIGINT Communications

Unknown

Run Date: 01/12/2005

(U) Last month, *SID today*'s DNI columnist Raul posted an [article on repairing damaged PDF files](#). We heard from a few analysts via the feedback mechanism that is available at the end of every article on *SID today*. What follows is a sampling of that feedback.

- (U//FOUO) *There is a program that tries to "repair" a PDF file. It was published as an Informal Technical Note (ITN) [S3T/TECH/11/2004 - S-249, 497 entitled "Repairing Broken PDF Files."](#)*
- (U)...*Another resource for extracting data has been CSE. I have obtained Perl routines that have enabled me to extract information from PDF files using the PDF structure.*
- (U//FOUO)...*A comment is made in this article that NSA does not have a capability against damaged PDF files. Actually, on the DESERTFOX (go DESERTFOX) team, we have a program called KEYSTONE that contains a damaged-PDF converter...This software only handles PDF files compressed using flate. Since flate is found in the vast majority of PDF files that we see, this has been a good solution for us for the past year. Sadly, it does not handle foreign character sets due to the tremendous amount of work required to pull out the (potentially damaged) code pages. The damaged-PDF converter has been ported to SORCERER and could even be provided as a stand-alone program. I sincerely hope that some effort was made to locate a program such as KEYSTONE before making the overarching statements, "No consideration has been given..." and "We should be able to deal with broken or damaged PDFs with great ease."*
- (S) *Why not just use pdftotext with UTF-8 encoding specified for output? This handles the decompression and encoding normalization issues, from what I've seen...(NOTE: pdftotext is one of the many valuable command-line utilities that comes with the xpdf application.)*

"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 ([DL sid comms](#))."
